The Effects of N-gram Probabilistic Measures on the Recognition and Production of Four-word Sequences.

Antoine Tremblay[1] and Benjamin V. Tucker[2]

[1]IWK Health Center

[2]University of Alberta

Address for correspondence:

Antoine Tremblay, Interdisciplinary Research, Goldbloom Research Pavilion, IWK Health Centre, 5850/5980 University Avenue, P.O. Box 9700, Halifax, Nova Scotia, B3K 6R8.

Email: trea26@gmail.com

Abstract

The present study investigates the processing and production of four-word sequences such as *I don't really know*, *at the age of*, and *I think it's the*. Specifically, we investigate the influence of families of probabilistic measures such as unigram, bigram, trigram, and quadgram frequency of occurrence, logarithmic (log) probability of occurrence, and mutual information. Log probability of occurrence emerged as the predominant predictor family in the onset latency analysis, suggesting that recognition is mainly underpinned by competition between a target N-gram and its family members. In contrast, the amount of experience one has with an N-gram (frequency of occurrence) surfaced as the most prominent predictor in production. Further, probabilistic measures tied to trigrams surfaced as the best predictors in the onset latency analysis, while the measures tied to unigrams were most predictive of production durations. finally, the interactions between probabilistic measures tied to unigrams, bigrams, trigrams, and quadgrams suggest that N-grams of different lengths are processed in parallel in both recognition and production.

Keywords: Multi-word sequences; N-grams; Speech processing; Speech production; Mixed-effects regression; Frequency of occurrence; Logit; Log probability of occurrence; Mutual information.

Three families of probabilistic measures have been investigated in speech production, namely, frequency of occurrence, logarithmic (log) probability of occurrence, and mutual information (e.g., 2001; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Ellis & Simpson-Vlach, 2009; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999). Frequency of occurrence is a widely used lexical measure that indexes the amount of experience one has with a given linguistic unit (Taft, 1979). While frequency measures a unit's prevalence relative to a complete set of units, probability of occurrence reflects the probability that a given unit will occur given one or more previously observed units. Related to these measures, mutual information indexes how strongly units depend on one another and how likely they will co-occur, whether frequent or not.

No study, as far as we know, has pitted frequency, probability, and mutual information against each other to determine which, if any, is a better predictor of four-word sequence production onset latencies and production durations. This is probably due to the fact that these measures are almost always highly correlated. When two or more correlated variables are present in a model, the true predictive power of either of them cannot be properly assessed (see section Reducing Collinearity). As a consequence, the majority of studies have investigates only one or two of these predictor types. The present study takes a first step at investigating interactions between these three variable families and how they affect laboratory recorded speech from lexical retrieval to production proper.

Ellis and Simpson-Vlach (2009), for instance, considered mutual information as well as frequency of occurrence, but they did not include conditional probability, nor did they account for the collinearity between frequency of occurrence and mutual information. Indeed, because they have not accounted for the collinearity between these variables, it is possible that mutual information could be substituted by frequency of occurrence in their analysis without changing the results. Such a state of affairs occurred in the analysis of word duration carried out by Gregory et al. (1999), where mutual information and bigram reverse conditional probability were inter-changeable.[1] Although we cannot determine the amount of collinearity between frequency and mutual information in the study by Ellis and Simpson-Vlach (2009), we would expect it to be very high given that in our data, collinearity between trigram and quadgram frequencies and mutual information values is relatively high (κ = 29, see below for discussion).

Bell et al. (2009) excluded mutual information from their analysis to circumvent problems tied to collinearity. Nevertheless, they left unchecked the potential problem of collinearity between word frequency (their prior probability) and conditional probabilities, which may also have been very high. In our data, the frequency of the second word of a sequence, FreqB, has a (high) correlation of 0.72 with the frequency of the first bigram, FreqAB.[2] Given that, for example, FreqAB is part of the equation used to compute word B’s backward conditional probability (i.e., FreqAB / FreqA, which has a correlation of 0.70 with FreqB and 0.55 with FreqAB), as a result the estimation of the amount of variability explained by this latter variable independently of others will be inaccurate. One study that pitted N-gram frequencies, probabilities, and mutual information against each other is Gregory et al. (1999). They investigated the effects of target word frequency, bigram and trigram frequency, probability of occurrence, and bigram mutual information on /t/ and /d/ deletions and flapping as well as word duration. Although they acknowledge that collinearity between these measures is problematic, they did not address this problem.

Interactions between (N-gram) probabilistic measures are very seldom taken into consideration, although they have something to say about the cognitive processes underlying language use. Baayen, Kuperman, and Bertram (2010) investigated interactions between frequency measures using data from word naming, visual lexical decision, and eye-tracking studies. They found second-order interactions between compound frequency, modifier frequency, and modifier family size, suggesting that multiple sources of linguistic information are processed in parallel (Baayen et al., 2010). Given their results, we expect to find similar interactions between probabilistic measures tied to unigrams, bigrams, trigrams, and quadgrams. At the level of the quadgram, interactions between quadgram and smaller N-gram probabilistic information would be similar to the results found by Baayen et al. (2010) for compound frequency and modifier frequency. We use an orthographic speech production task to investigate the following research questions: (1) Whether the frequency of occurrence, log probability of occurrence, and mutual information of unigrams, bigrams, trigrams, and quadgrams (N-grams) affect onset latencies and production durations in laboratory recorded speech; (2) Which one of these N-grams and probabilistic measures are better predictors of (a) onset latencies and (b) production durations; and (3) Whether there are any second-order (linear) interactions between these probabilistic measures. We hypothesize that we will find shorter voice onset latencies and sequence durations for N-grams that are more predictable and higher frequency as well as second-order interactions between probabilistic variables tied to N-grams of different lengths.

Experiment

Participants

This experiment examines the productions and response latencies from 17 young adult undergraduate students who are native speakers of English (7 males and 10 females). All participants were recruited from the University of Alberta community and were paid for their participation.

Material

The most frequent quadgrams from the *Phrases in English* website (Fletcher, 2008), which incorporates data from the BYU-BNC (Davies, 2004) version of the *The British National Corpus*. This amounted to 112 four-word sequences ranging in frequency from 117 to 12 per million words. We then randomly selected 320 quadgrams again from the *Phrases in English* website; quadgrams frequencies ranged from 11 to 0.3 per million words. Overall, our stimulus list comprised 432 four-word sequences, with frequencies approximating a normal distributed. We subsequently extracted frequency counts for each item of our stimulus list from the *The Corpus of Contemporary American English* (COCA: Davies, 2008). Contractions such as *don't, you've,* and *wasn't* were treated as one word.

Procedure Participants were seated in an Industrial Acoustics Corporation sound attenuated booth with a computer monitor placed on the outside of a window facing the participant. Each four-word string was preceded by a fixation cross (font: Arial; size: 36; position: center) for 500 ms followed by 20 ms transition between the fixation and the appearance of the target string. Participants were then visually presented with one of the 432 four-word sequences (font: Arial; size: 36; position: center) for 1500 ms and asked to produce the sequence as quickly as possible after it appeared on the screen. Sequences were randomized and presented with an interstmulus interval of 1000 ms. Participants were given the opportunity to take two short breaks during the experiment, though many opted to continue without a break. Two microphones were situated in close proximity to the speaker's mouth. The first one was used as a voice key for the elicitation of response (onset) latencies, that is, the time from the appearance of a sequence on the screen to the time a participant began to produce it. The second microphone recorded the speech of each speaker for later analysis of sequence duration and notation of production errors.

Analyzing the Data – Preliminary Considerations

Before moving on to the main analyses, we discuss the problem of reducing collinearity, iteratively fitting linear mixed-effects models, and interpreting these results.

Reducing Collinearity

One assumption of regression is that the predictor variables are mutually independent. This ensures that a one unit increase in variable X has effect W on the dependent variable when other predictor variables are kept constant. If, for instance, variable X is highly correlated with variable Z, one will be unable to ascertain whether effect W is attributable to X or Z given that a one unit increase in X will forcibly be accompanied by a similar increase/decrease in Z. In addition to making the interpretation of the estimates of the regression coefficients difficult, it can inflate the standard deviations of these estimates, thus decreasing statistical power (Glantz & Slinker, 1990), hinder the process of selecting truly important variables (Harrell, 2001, pp. 64–65), and render certain mathematical operations impossible or unstable (Kline, 2005, pp. 56–57). A number of measures are available to determine whether predictors are collinear: (1) squared multiple correlation, R_{smc}, between each variable and all the rest, where a *Rsmc* greater than 0.90 suggests collinearity (Kline, 2005, p. 57); (2) tolerance, $1 - R_{smc}$, indicating the proportion of total standardized variance that is unique, where a tolerance value below 0.10 indicates collinearity; (3) variance inflation factor (*VIF*), $1/(1 - R_{smc})$, which is the ratio of the total standardized variance to unique variance, where a *VIF* greater than 10 indicates that a predictor is redundant; (4) pairwise Spearman correlations, where a correlation of 0.30 indicates collinearity between two predictors and a correlation of 0.90 indicates that the two variables are redundant (Kline, 2005, p. 56); (5) condition number, κ (Belsley, Kuh, & Welsch, 1980), which gives an overall index of the amount of collinearity within a set of predictors. A condition number between 1 and 10 indicates that there is no collinearity while $\kappa > 10$ indicates collinearity, which is considered high above 30 (Belsley et al., 1980; Baayen, 2008).[3]

We will prefer the κ index given that “indexes such as *VIF* are not very informative as some variables are algebraically connected to each other”, which is the case here, and “summarizing collinear variables using a summary score is more powerful and stable than arbitrary selection of one variable in a group of collinear variables” (Harrell, 2001, p. 65).

In the present study we find massive collinearity between predictors (κ = 5.332e+16) and the assumption of independence is patently broken. By way of example, the correlation between FreqB and FreqAB is 0.72, meaning that when FreqB increases from 1 to 2, FreqAB will also increase from 1 to (roughly) 2. Thus it would not be possible to keep FreqB constant and vary FreqAB to determine the independent (partial) effect of this latter variable on onset latencies and sequence durations.

The simplest way to reduce collinearity is to center our variables (Cronbach, 1987). Doing so reduces κ to 1.813e+16, that is, by a factor of 2.9, which remains too high.[4] Further steps that can be taken include eliminating variables (Kline, 2005). This option, however, runs counter to the goals of the paper. Yet another option is to combine highly correlated predictors into a composite variable (Harrell, 2001; Kline, 2005; Baayen, 2008). For instance, we could perform a principal components analysis (PCA) on our set of predictor variables to create, for instance, five new non-collinear composite variables. Reducing dimensionality in such a manner, however, also runs counter to the goals of the paper. The last option involves residualization. In essence, residualization is a method to statistically control for the influence of, for example, variables *V2* , *V3* , . . . , *Vn* on variable *V1*. Residualization creates a new variable by taking the residuals of a linear model where the dependent variable is the to-be-residualized predictor and the independent variables are the collinear predictors. Let us consider, for example, the creation of the variable $FreqAB_{residualized}$, which is illustrated in figure 1 with the help of a Venn diagram.

[figure 1 about here.]

The light grey circles represent FreqA and FreqB and the dark grey one is FreqAB. The degree of correlation between these variables is depicted by the amount of overlap between the circles. The linear regression FreqAB as a function of FreqA + FreqB removes the portions of FreqA and FreqB that overlap with FreqAB and the new variable $FreqAB_{residualized}$ is the portion of the dark grey circle that does not overlap with any predictor.[5] For each N-gram measure, we created a new residualized variable where the independent variables in the linear model were the lexical measures of every sub-chunk contained within the N-gram. Note that we performed this procedure only within variable families (i.e., frequencies, log probabilities, and mutual information) and that single-word frequencies were not residualized. Table 1 shows the correlations between the mean centered variables and the residualized ones. Note that in general, the correlations decrease as the length of the N-gram increases. This is to be expected given that more is “taken out of” longer N-gram probabilistic measures. Compare the linear model used to residualize FreqAB mentioned above to the one used to residualize quadgram frequency of occurrence: FreqABCD as a function of FreqA + FreqB + FreqC + FreqAB + FreqBC + FreqCD +

FreqABC + FreqBCD.

[Table 1 about here.]

Although the condition number has now decreased to 6.34e+15 (a 2.9 fold decrease on top of the previous one), it is still too high, thus we will be unable to include all these predictors in one single model. A workaround is to first back-fit one model for each predictor family (i.e., one for frequencies, one for log probabilities, and one for mutual information), then take, for each model, the predictors that survived the back-fitting process and conjoin them in a fourth model, which will also be back-fitted. Such an approach is acceptable given that centered-residualized predictors within each family have low collinearity (Frequency: κ_{raw} = 60.97, $\kappa_{centered}$ = 10.05, $\kappa_{residualized}$ = 2.24; Logit: κ_{raw} = 13.85, $\kappa_{centered}$ = 8.99, $\kappa_{residualized}$ = 2.23; Mutual information: κ_{raw} = 35.93, $\kappa_{centered}$ = 13, $\kappa_{residualized}$ = 2.24).

Iterative Model fitting

The general iterative model fitting procedure used in this paper is as follows: (1) fitting of an initial linear mixed-effects regression model; (2) Removal of data points with undue influence on the regression (outliers); (3) Re-fitting of the model to the trimmed data; and (4) Back-fitting of the model.

We start by fitting an initial linear mixed-effects regression model (LMER) to the un-trimmed data with by-subject and by-item random intercepts in R version 2.10.1 (R Development Core Team, 2009) using the *lmer* function from package *lme4* (Bates, Maechler, & Bolker, 2011).[6] Then, outliers 2.5 standard deviations above and below the model residuals mean are removed (de Vaann, Schreuder, & Baayen, 2007; Baayen, 2008).

After re-fitting the model to the trimmed data, it is back-fitted using the following process (implemented as function *fitLMER.fnc* in package *LMERConvenienceFunctions*; Tremblay, 2011): (1) first, highest-order interaction terms are considered. (a) The model term with the lowest t-value below 2 is identified. (b) A new model without this term is fitted. (c) The more complex and simpler models are compared by way of a log-likelihood ratio test (LLRT; Pinheiro & Bates, 2000, p. 83–87). If the result of the LLRT indicates that the term under consideration does not increase model fit (i.e., $p > 0.05$), it is removed; otherwise it is kept. (d) The process moves on to the next model term with the smallest t-value below 2 and steps (a)–(c) are repeated. (2) Once all highest-order interaction terms have been evaluated, the process moves down to the second highest order interactions and step (1) is repeated with the following addition: If a term would be removed from the model, but it is part of a high-order interaction, it is kept in the model (i.e., the marginality principle; Venables, 1998). Once all terms of the interaction level have been evaluated, the process moves down to the next lower-order level until main effects have been evaluated. (3) The random effects structure is forward-fitted. (a) For each fixed effect that survived the backward fitting process, a new (more complex) model is fitted that includes by-subject random slopes for that fixed-effect. (b) The more complex and simpler models are compared by way of LLRT. If the test is significant, the random effects term is kept; otherwise it is dropped. (4) The model is subsequently back-fitted once more. This is done given that the inclusion of certain random effects sometimes renders certain fixed effects non-significant. (5) finally, for each continuous predictor that survived the back-fitting process, values 2.5 standard deviations above and below the predictor mean are removed and the model is back-fitted again. This final step is an effort to diminish the potential undue influence of extreme predictor values on the regression model. The fact that the estimated coefficient of a predictor is no longer significant after the removal of its extreme values would indicate that the effect associated with it was driven by those extreme values (it is thus dropped from the model).

Interpreting the Results

A linear (mixed-effects) regression model allows one to determine linear relationships between two or more predictors. Mixed-effects models can be expressed, in simple mathematical terms, as

$$y = Var1\ \beta_1 + Var_2\ \beta_2 + \cdots + Var_n\ \beta_n + b + \varepsilon,$$

where y is the observed data, Var_1 , Var_2 , . . . , Var_n are predictor variables, β_1, β_2, . . . , β_n are coefficients weighing these variables, b represents variability tied to subjects and/or items (i.e., random effects), and ε is the residual error. In other words, the observed data is equal to the weighted sum of the covariates, plus subject and/or item variability, plus random error. If there is no collinearity between predictors, then the beta coefficients give, for the average unknown subject and item, an indication of the expected change in y given a one unit increase in a predictor with all others held constant. For instance, $\beta_{FreqABCD}$ would be equal to -2.5 if an increase of 1 in quadgram frequency would be associated with a 2.5 millisecond decrease in onset latency while keeping all other N-gram frequencies constant.

In sum, we believe that the model resulting from the process described above is the one that is, of the set of possible statistical models, (1) the most well-formed in terms of meeting the underlying regression assumptions; (2) the most parsimonious in that it is the model with as few parameters as possible; and (3) the most stable and generalizable given that (a) spurious effects driven by unduly influential extreme values are dealt with both at the data- and predictor-level, and (b) by-subject and by-item variability with regards to the intercepts and slopes is taken into account.

Data Analysis

The raw data contained 7344 data points (17 participants x 432 items four-word sequences). One data point corresponds to one onset latency and one sequence duration value (in milliseconds). Before we analyzed the data, we excluded five items with a frequency of 0 in coca (85 data points or 1.2%) as well as one participant who had a substantial amount of missing data and production errors (39 data points or 0.5%). In some cases, a response was not triggered either because the participant was too far away from the microphone or spoke too softly; these data points were also excluded (869 data points or 11.8%). finally, productions of each sequence were analyzed using *Praat* (Boersma & Weenink, 2010). Research assistants noted errors and measured the sequence duration. Several types of production errors and notes were indicated for each sequence: Deleted segments (*a few of the* → *a few o the*), missing words (f*ew of the*), incorrect word choice (*a few and the*), and repetitions (*a a few of the*). Deleted segments were not considered errors but rather natural occurrences of speech production and were expected for higher frequency sequences; they were thus retained while all other instances were excluded from the main analysis below (1258 data points or 17.1%). Overall, 2251 data points were removed (30.7% of the data).

Dependent and Independent Variables

Two dependent variables were analyzed: "onset latency" and "sequence duration" reflecting recognition and production respectively. We also considered a number of independent variables, which are briefly described below. A log transform (natural log) was used for dependent and numerically independent variables to normalize their distribution.

Frequency of Occurrence (Freq). Frequency of occurrence is a widely used lexical measure that indexes the amount of experience a speaker has with, for instance, a given word, in the number of times it occurs within a set of words such as the lexicon (e.g., the quadgram at the age of has a frequency of 9.2 per million words, which is relatively frequent). We considered in our analyses single word frequencies (FreqA, FreqB, FreqC, FreqD), bigram frequencies (FreqAB, FreqBC, FreqCD), trigram frequencies (FreqABC, FreqBCD), and quadgram frequency (FreqABCD), where the capital letters A, B, C, and D stand for single-word positions within a four-word sequence. We standardized our frequency measures to per-million words; there were approximately 385 million words in coca at the time frequency counts were extracted.

Log Probability of Occurrence (Logit). Logits, or log probability of occurrences (Tremblay, 2009; Tremblay & Baayen, 2010), provide an index of the probability of a word occurring given a certain context (e.g., the occurring given at or of occurring given the sequence at the end). As such, logits are related to forward conditional probabilities (Gregory et al., 1999), predictability (Frisson, Rayner, & Pickering, 2005), and cloze probability (e.g. Wlotko & Federmeier, 2007). In this study, we considered bigram, trigram and quadgram logits. They were calculated by taking the log of the frequency of a sequence (e.g., *the end*) divided by the summed frequency of all the possible sequences that share the same "context" (e.g., *the beginning*, *the sea*, *the man*, ...) minus the frequency of the sequence plus 1 (to back away from dividing by 0). For example, LogitABCD was calculated as *log((FreqABCD/(FreqABC − FreqABCD)) + 1)*.

Logit can also be construed as indexing how frequent a particular item is relative to the other members of its "family". For example, the quadgram at the age of has a logit of 3.4, which indicates that it is much more frequent than the remaining 44 family members, which have a summed frequency of 0.3 per million words (the second and third most frequent members, at the age when and at the age where, each have a frequency of 0.08 and a logit of -1.4).[7] If, for instance, the target sequence for production is at the age of (logit of 3.4), as was the case in one of our trials, one could imagine that the rate of activation will be quicker for this sequence than for its competitors (i.e., the other members of the family at the age), which may lead to a faster production onset. If, however, the target were at the age when (logit of -1.4), it is conceivable that higher frequency competitors such as at the age of would reach activation threshold more quickly and may have to be inhibited, potentially resulting is in a slower sequence production onset.

Mutual Information (Mi). Mutual information is a commonly used measure that indexes how strongly words are associated to one another and how likely they are to co-occur (e.g., Gregory et al., 1999; Pluymaekers, Ernestus, & Baayen, 2005; Ellis & Simpson-Vlach, 2009). The higher the mutual information score, the greater the coherence/dependence between the words. This measure has the advantage of distinguishing low-frequency words that commonly occur together from sequences of high-frequency words that are unrelated. Mutual information scores were calculated by taking the log of the probability of occurrence of a sequence divided by the product of the frequencies of the single-words that compose the sequence. For example, the mutual information score of the whole sequence, *MiABCD, as log(P(ABCD) / (P (A) x P (B) x P (C) x P (D))).*

Other Independent Variables. The remaining independent variables we considered in our analyses are listed and briefly described in Table 2.

[Table 2 about here.]

The Onset Latency and Production Duration Models

Each model was fitted as outlined in section *Iterative Model fitting* above. Each one of the initial six models included single word frequencies (FreqA, FreqB, FreqC, and FreqD) as well as $\text{Length}_{\text{residualized}}$, NumSyll, Manner, PhraseABCD, Trial, and PrevTrialPC1 in addition to the lexical variables of one of the three predictor families. The models included every possible two-way interaction as well as by-subject and by-item random intercepts. We then took the terms that survived the back-fitting process and conjoined them into a fourth model, which was also back-fitted using the same process. Note that each one of the eight models was initially fitted to the un-trimmed data. Outliers were subsequently removed, which represented in each case approximately 2% of the data.

Regarding the onset latency analysis more specifically, the surviving predictors had a high degree of collinearity ($\kappa = 31$). Much of the collinearity was due to a few highly correlated pairs of variables having a correlation greater than 0.65, namely FreqAB and LogitAB, FreqBC and LogitBC, LogitABC and MiABC, LogitBCD and MiBCD, as well as FreqABCD and LogitABCD. This meant that these variables had similar predictive power in the model. We thus conjoined these variable pairs by performing a principal components analysis on each pair and taking the first principal component, which accounted for more than 95% of the within-pair variability (Baayen, 2008). The old variables were then replaced by the new ones PC[FreqAB / LogitAB], PC[FreqBC / LogitBC], PC[LogitABC / MiABC], PC[LogitBCD / MiBCD], and PC[FreqABCD / LogitABCD]. Collinearity between predictors was now low (κ = 2.7). We once more back-fitted the new model, which resulted in a few predictors and interactions being removed.

Turning to the sequence duration analysis, collinearity between predictors was acceptable (κ = 10.5). Nonetheless, two variable pairs had a high correlation (R > 0.70), namely FreqABC and LogitABC as well as FreqCD and LogitCD. These two pairs were conjoined by way of principal components analysis into the new variables PC[FreqAB / LogitAB] and PC[FreqCD / LogitCD]; collinearity between predictors was now lower (κ = 6.1). We back-fitted the model once more. finally, we removed, for each surviving predictor, items with extreme values. A total of 334 items remained in the onset latency analysis and 298 in the sequence duration analysis. Model criticism plots indicated that the residuals were approximately normally distributed with a constant variance, and that no data point unduly influenced the regressions. Probability values (i.e., *p*-values) and 95% confidence intervals were calculated by way of Markov Chain Monte Carlo simulation (MCMC) using the *pvals.fnc* function from package *languageR* (Baayen, 2011).

Results and Discussion

The result of our analysis are summarized in Tables 3, 4, 5, and 6 and visually in figures 2 and 3. Each figure contains several panels, which show the effects of significant predictors on onset latencies (figure 2) and sequence durations (figure 3). The grey lines at the bottom of the continuous main effect panels, right above the x -axis, represent the distribution of a predictor. In panels depicting interactions between two continuous variables, the distribution is represented by the grey lines drawn between the solid black line and the lightest broken grey line, which effectively shades this area of the plot. Take, for example, panel I of figure 2, which graphs the PC[FreqABCD / LogitABCD] X FreqB interaction. The solid black line shows the effect of PC[FreqABCD / LogitABCD] on onset latencies when the frequency of the second word of a sequence (FreqB) is set at its first quantile (i.e., at its lowest) and the dashed grey lines represent the effect of PC[FreqABCD / LogitABCD] when FreqB is set to one of the remaining quantiles (i.e., 25th, 50th, 75th, and 100th).

Onset Latency

Unigram frequencies as well as trigram logits and mutual information values were the predominant variables affecting onset latencies. Logits and mutual information values are interchangeable at the level of the trigram, as shown by the presence of the two variables PC[LogitABC / MiABC] and PC[LogitBCD / MiBCD] created from both variables families. That is, a model including LogitABC and LogitBCD would have approximately the same predictive power as one including MiABC and MiBCD instead. The presence of the variables PC[FreqAB / LogitAB] and PC[FreqABCD / LogitABCD] in the final model indicates that, at the bigram and quadgram level, frequencies and logits are also interchangeable but have a smaller effect on onset latencies than trigrams.

Interestingly, the frequency of occurrence of the second word of a sequence, FreqB, interacted with probabilistic variables tied to bigrams (PC[FreqAB/LogitAB]), trigrams (PC[LogitBCD/MiBCD]), and quadgrams (PC[FreqABD/LogitABCD]). Why is it that FreqB, and not FreqA, FreqC, or FreqD interacting with so many variables?

[Table 3 about here.]

[Table 4 about here.]

[figure 2 about here.]

The second word of a sequence appeared, more often than not, in the center of the screen, roughly where the fixation cross had appeared immediately before. A $\chi 2$ test between the second and third words, where 50000 replications were used in a Monte Carlo simulation, revealed that word B appeared in the center of the screen significantly more often than word C did (second word, B: 141/334, third word, C: 95/334, $\chi 2 = 8.99$, simulated $p = 0.004$).[8] Given that the participants' attention was primarily focused at the very position where the second word of a sequence appeared, it is not surprising that its frequency played a major role in lexical access and production onset, interacting with other lexical variables. The third word of a sequence also appeared at roughly the same position the fixation cross had previously been presented, albeit to a lesser extent, and it is quite possible that the frequency of the third word also affected onset latencies in these cases.

However, its effect may not have emerged in our analyses given the relative rarity with which it occurred. It is also possible that the frequency of word B affected onset latencies rather than word C because there were more content words in the second position than in the third. Indeed, content words are known to be fixated more than twice as often as non-content words (Rayner, 1998, p. 375, and references cited therein). In eleven cases, both words B and C were content words (e.g., *the same way as*, *to take account of*, *a large number of*) and in 108 sequences these two words were non-content words (e.g., *and I don't like*, *that's what you want*, *it would have been*). In 100 sequences, word B was a content word followed by a non-content one (e.g., *the first of these, I think it's very, a result of the*), and in 115 sequences word B was a non-content word followed by a content word (e.g., *I don't like that*, *it doesn't matter what*, *it is difficult to*). A $\chi 2$ test indicates that there is an equal number of content words in the second and third positions of a sequence ($\chi 2$ = 1.05, simulated p = 0.34 based on 50000 replications) meaning that it is unlikely that this affected the responses. The level of lexical activation of the second word of a sequence may thus have benefited from having been the first portion of the sequence to impinge on a participant's visual system. Consequently, activation of word B would have begun to increase as a function of its frequency before any of the other portions of the carrier sequence. Although the increase of WordB's level of activation as a function of FreqB was beneficial for the first bigram (PC[FreqAB / LogitAB] became more facilitatory as FreqB increased), competition was engaged between word B and the second trigram, BCD, as as well as the whole sequence, ABCD: The higher the level of activation of these two units (as indexed by their predictor values), the more time it took to resolve the competition resulting in longer production onsets.

Given the high correlation between the frequency and grammatical category of a word (function words are higher frequency than content words), we might expect that FreqB can be replaced by the categorical variable WordTypeB with levels “content word” and “function word”.[9] Nevertheless, none of the interactions with PC[FreqAB / LogitAB], PC[LogitBCD / MiBCD], and PC[LogitBCD / MiBCD] nor the main effect of WordTypeB reached significance. We also considered the addition of the three-way interactions WordTypeB X FreqB X PC[FreqAB / LogitAB], WordTypeB X FreqB X PC[LogitBCD / MiBCD], and WordTypeB X FreqB X PC[LogitBCD / MiBCD], as well as lower-order interactions to the model. While the three interactions involving FreqB remained significant, the only significant effect involving WordTypeB was a FreqB X WordTypeB interaction ($F(1, 4860) = 4.1$, $p = 0.04$; MCMC $\beta = -0.01$, $t = -2.0$, MCMC $p = 0.03$, 95% confidence interval = −0.02 to −0.0006). Which shows an inhibitory effect of FreqB on onset latencies was greater for content words than function words.

Production Duration

The results from the production duration analysis (summarized in Tables 5 and 6 and figure 3) were quite different from the ones found in the onset latency analysis. There were many more main effects and interactions in the former analysis, where unigram frequencies made up the largest proportion of the effects. Unigram frequencies did not interact with other probabilistic measures tied to larger N-grams. The number of trigram and quadgram probabilistic effects was also considerable, albeit to a lesser degree. finally, frequencies of occurrence appear to have been the most important of the three variable families, which were, in a few instances, interchangeable with logits.

[Table 5 about here.]

[Table 6 about here.]

[figure 3 about here.]

General Discussion

In this paper we set out to determine whether frequency of occurrence, log probability of occurrence, and mutual information of N-grams separately affect onset latencies and production durations in laboratory recorded speech. The results of the two analyses reported here indicate that these different measures indeed have a separate effect, albeit they do so differently.

We also endeavored to determine which family of probabilistic measures better predicts onset latencies and production durations. The percentage of deviance explained by frequencies of occurrence, log probabilities of occurrence, and mutual information values in each of the onset latency and production duration analyses is provided in Table 7.

[Table 7 about here.]

These values were obtained by summing the deviance explained for each model term in which a probabilistic measure was involved. Note that unigram frequencies were included in the calculation of the value for the frequency of occurrence family. In the onset latency analysis, log probabilities of occurrence were the most important variable family (0.93%), closely followed by mutual information values (0.85%), and finally by frequencies of occurrence (0.21%). This suggests that the main process underlying recognition is one of competition between N-grams and their family members (Marslen-Wilson, 1995) and that the degree of uniqueness of an N-gram, whether frequently occurring or not, is less important. This provides evidence for a secondary (potentially simultaneous) process possibly involving the holistic retrieval of N-grams.

On the contrary, frequencies of occurrence were by far the most important family of probabilistic measures affecting production (1.11%). Log probabilities of occurrence accounted for a minimal amount of variance (0.05%) and, surprisingly, mutual information values did not have any predictive power. This may indicate that in production the number of times one has accessed/produced a linguistic item is important. Thus, the neuromotor routines that instantiate a sequence's phonetic form become more fluent with repetition resulting in reduction and coarticulation (Bybee, 2001, 2006, and references cited therein).

Although N-gram probabilistic measures up to the quadgram affected onset latencies and production durations, which type of N-gram was the most important? Table 8 lists the amount of deviance explained by unigrams, bigrams, trigrams, and quadgrams.

[Table 8 about here.]

It is apparent from Table 8 that probabilistic information tied to trigrams was the most predictive of onset latencies (0.29%) closely followed by unigram frequencies (0.25%), which is most probably due to the fact that a participant's focus of attention was more often than not initially directed to the second word of a quadgram. We are uncertain what the implications of this are regarding the recognition of multi-word sequences. Three-word sequences may be the most efficient unit of processing in that they would enable the linguistic system to strike a balance between keeping the longest possible N-gram (in number of words) in short-term memory and the amount of cognitive resources needed to keep it there.

In contrast, probabilistic measures tied to trigrams had very little predictive power with regards to production durations (0.03%). Although bigram and quadgram variables were relatively important (0.17% and 0.12% respectively), unigram frequencies were by far the most predictive (0.96%). It may be that individual words are a fundamental organizational unit of speech production and thus frequency plays an important role during production.

Finally, the presence of interactions between N-gram probabilistic measures are not in line with types of models of visual recognition and speech production that would assume that four-word sequences are (de)composed in stages either from unigrams up to quadgrams or the other way round. Rather, the results reported here support the idea of a dynamic linguistic system that uses, in parallel, multiple sources of probabilistic information at different supra-lexical levels of structure.

Conclusion

The results reported here illustrate how complex and dynamic the linguistic system is. Log probability of occurrence emerged as the predominant predictor family in the onset latency analysis, suggesting that recognition is mainly underpinned by a mechanism whereby a target N-gram competes with its family members. In contrast, the amount of experience one has with an N-gram (frequency) surfaced as the most important predictor family in the analysis of production duration. Somewhat surprisingly, the cohesiveness of an N-gram (mutual information) played only a minor role in recognition and none at all in production. Although unigrams, bigrams, trigrams, and quadgrams all affected both recognition and production, trigrams arose as the most important N-gram in the former stage, whereas unigrams were the most important one in the latter stage. finally, the finding that probabilistic measures tied to N-grams up to four-words long interacted with each other in the onset latency and production duration analyses suggests that they are processed in parallel in both recognition and production.

References

Allen, M. (1997). *Understanding Regression Analysis*. New York: Plenum Press.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*, 67–82.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, U.K.: Cambridge University Press.

Baayen, R. H. (2011). *languageR: Data Sets and Functions with "Analyzing Linguistic Data: A Practical Introduction to Statistics"*. [Computer software manual]. Available from http://CRAN.R-project.org/package=languageR (R package version 1.2)

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Baayen, R. H., Kuperman, V., & Bertram, R. (2010). Frequency effects in compound processing. In S. Scalise & I. Vogel (Eds.), *Compounding*. (pp. 257–270). Amsterdam and Philadelphia: John Benjamins.

Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words. a regression study across tasks and modalities. *The Mental Lexicon, 2*, 419–463.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear Mixed-effects Models using S4 Classes* [Computer software manual]. Available from http://CRAN.R-project.org/package=lme4 (R package version 0.999375-39)

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 92–111.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics. Identifying influential data and sources of collinearity.* New York: Wiley.

Boersma, P., & Weenink, D. (2010). *Praat: Doing Phonetics by Computer (computer program)*. http://www.praat.org/, Version 5.1.44.

Bybee, J. (2001). P*honology and Language Use*. Cambridge: Cambridge University Press.

Bybee, J. (2006). From usage to grammar: The minds response to repetition. *Language, 82*, 711–733.

Cronbach, L. J. (1987). Statistical test for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin, 102*, 414–417.

Davies, M. (2004). *BYU-BNC: The British National Corpus*. Available on-line at http://corpus.byu.edu/bnc.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 400+ Million Words, 1990-Present*. Available on-line at http://www.americancorpus.org.

de Vaan, L., Schreuder, R., & Baayen, R. H. (2007). Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon, 2,* 1–24.

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers:

Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory, 5*, 61–78.

Fletcher, W. H. (2008). *Phrases in English.* Available on-line at http://pie.usna.edu/index.html.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 862–877.

Gelman, A., & Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press.

Glantz, S., & Slinker, B. (1990). *Primer of Applied Regression and analysis of Variance*. New York: McGraw-Hill.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *CLS–99 , 35*, 151–166.

Harrell, F. (2001). *Regression Modeling Strategies*. Berlin: Springer.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (p. 229-254). Amsterdam: Benjamins.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling. Second Edition*. New York and London: The Guilford Press.

Marslen-Wilson, W. (1995). Activation, competition, and frequency in lexical access. In G. T. Altman (Ed.), Cognitive Models of Speech Processing. (pp. 148–172). Cambridge, MA: MIT Press.

Newman, A., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. (Accepted). The influence of language proficiency on lexical-semantic processing in native and late learners of English: Linear mixed effects modeling of ERP and proficiency data. *Journal of Cognitive Neuroscience*.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken dutch. *Phonetica, 62*, 146–159.

R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing* [Computer software manual]. Vienna, Austria. Available from http://www.R-project.org (ISBN 3-900051-07-0)

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Scheibman, J., & Bybee, J. (1999). The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics, 37*, 575-596.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition, 7*, 263–272.

Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology-learning Memory and Cognition*, *27*, 117–138.

Tremblay, A. (2009). Processing Advantages of Lexical Bundles: Evidence from Self-paced Reading, Word ad Sentence Recall, and Free Recall with Event-Related Brain Potential Recordings. PhD dissertation, University of Alberta, Edmonton, Canada.

Tremblay, A. (2011). *LMERConvenienceFunctions: A Suite of Functions to Back-fit Fixed Effects and Forward-fit Random Effects*. [Computer software manual]. Available from http://CRAN.Rproject.org/package=LMERConvenienceFunctions (R package version 1.5)

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (p. 151-173). London and New York: Continuum.

Venables, W. (1998). *Exegeses on linear models*. Available from http://www.stats.ox.ac.uk/pub/MASS3/Exegeses.pdf. (Paper presented to the S-PLUS User's Conference, Washington, DC, 8–9th October, 1998).

Wlotko, E. W., & Federmeier, K. D. (2007). finding the right word: Hemispheric asymmetries in the use of sentence context information. *Neuropsychologia*, *45*, 3001–3014.

Wu, H., & Zhang, J. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. Hoboken, NJ: Wiley.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60*, 502–529.

Footnotes

[1] "As in the case of deletion, mutual information could be replaced in the model in [their] Table 3 with reverse conditional bigram probability without changing the predictive capacity of the model", (Gregory et al., 1999, p. 15).

[2] The capital letters A and B stand for the first and the second word of a four-word sequence, respectively.

[3] The condition number, κ, is equal to highest covariance of set of variables divided by the lowest one, where covariances are obtained by taking the diagonal of a singular value decomposition of that set.

[4] Mean centering reduces the covariance between a set of variables and concomitantly the condition number.

[5] See, e.g., Newman, Tremblay, Nichols, Neville, and Ullman (Accepted) and Arnon and Snider (2010) for a similar approach. Also see Allen (1997) for more details on residualization.

[6] LMER is a natural tool for modeling repeated measures (Wu & Zhang, 2006). Details about mixed-effects modeling can be found in a number of recent papers and books (e.g. Pinheiro & Bates, 2000; Wu & Zhang, 2006; Gelman & Hill, 2007; Baayen, 2008; Baayen, Davidson, & Bates, 2008).

[7] Positive log probability values indicate that the target N-gram is more frequent than the summed frequency of the remaining family members.

[8] The fixation cross appeared in the middle of the second and the third word in 98/334 cases.

[9] $M_{FreqB\ function}$ = 5322 per million; $M_{FreqB\ content}$ = 418 per million, $R_{FreqB\ -\ WordTypeB}$ = 0.63, $F(1, 174) = 117.8$, $p < 0.0001$.

Tables

Table 1: *Correlations between Mean-centered and Residualized Variables*

	AB	BC	CD	ABC	BCD	ABCD
Frequency of Occurrence	0.55	0.63	0.48	0.49	0.48	0.34
Log Probability of Occurrence	0.47	0.57	0.65	0.59	0.58	0.37
mutual Information	0.77	0.85	0.81	0.38	0.43	0.22

Table 2: Other independent Variables Considered in the Analyses

Trial	Trial number in the experiment (from 1 to 432).
PrevTrialsPC1	The first principal component of the reaction times (i.e., onset latency or sequence duration) from the three previous trials (Taylor & Lupker, 2001; Baayen, Wurm, & Aycock, 2007; de Vaan et al., 2007).
Manner	Whether the first phoneme of a sequence is an approximant (39 sequences), a fricative (126 sequences), a nasal (8 sequences), a stop (62 sequences), or a vowel (192 sequences). This variable is known to affect measures of voice latency (Baayen et al., 2007; Yap & Balota, 2009).
Length	Length of a sequence in number of letters (7 to 29 letters long).
NumSyll	Length of a sequence in number of syllables (4 to 9 syllables long).
PhraseABCD	Whether the sequence is a phrase (117 sequences) or a non-phrase (310 sequences). Phrases can stand alone, (e.g., *end of the year*, *I don't really know*, *at the same time*, and *I have to say*) but non-phrases cannot (e.g., *it would be a*, *at the age of*, *this is not a*, *we've got to get*, and *I think it's the*).

Table 3: *Results of the Onset Latency Analysis – Part 1*

Variable	*F*	Num. *df*	*p*	X^2 LLRT	*df*	*p*	R^2
Trial	11	1	< 0.001	8.6	1	< 0.001	0.0011
Number of Syllables	1.8	1	0.18	4.3	1	0.04	0.0002
Manner of articulation	12.8	4	< 0.001	43.9	4	< 0.001	0.005
PhraseABCD	7.5	1	0.01	8.4	1	0.004	0.0007
FreqD	0.9	1	0.35	9.5	1	0.002	0.0001
PC[LogitABC / MiABC	20.3	1	< 0.001	20.8	1	< 0.001	0.002
FreqB X PC[FreqAB / LogitAB]	9.5	1	< 0.001	11.9	1	< 0.001	0.0009
FreqB X PC[LogitBCD / MiBCD]	8.8	1	< 0.001	9.1	1	0.003	0.0009
FreqB X PC[FreqABCD / LogitABCD]	6.6	1	0.01	10.8	1	0.005	0.0006

Notes. Denominator *df* = 4862. LLRT stands for log-likelihood ratio test.

Table 4: *Results of the Onset Latency Analysis – Part 2*

Variable	MCMC β	95% CI	*t*	MCMC *p*
Trial	0.0002	0.0001 to 0.0003	3.3	0.01
Number of Syllables	0.0105	0.0005 to 0.0194	2.1	0.03
Manner of articulation	--	--	--	--
PhraseABCD	-0.0221	-0.0104 to -0.0024	-3	< 0.001
FreqD	-0.0065	-0.0105 to -0.0026	-3	< 0.001
PC[LogitABC / MiABC	0.0109	0.0062 to 0.0151	4.5	< 0.001
FreqB X PC[FreqAB / LogitAB]	-0.0042	-0.0065 to -0.002	-3.4	< 0.001
FreqB X PC[LogitBCD / MiBCD]	0.0033	0.0013 to 0.0054	3	< 0.001
FreqB X PC[FreqABCD / LogitABCD]	0.0025	0.0004 to 0.0047	2.1	0.02

Notes. MCMC stands for Markov Chain Monte Carlo. *CI* stands for confidence intervals.

Table 5: *Results of the Production Duration Analysis – Part 1*

Variable	*F*	Num. *df*	*p*	X^2 LLRT	*df*	*p*	R^2
Number Syllables	187.5	1	< 0.001	149.2	1	< 0.001	0.0102
Manner of articulation	5.8	4	< 0.001	18.5	4	0.001	0.0013
FreqB X PhraseABCD	7.7	1	0.01	4.9	1	0.027	0.0004
FreqC	139.2	1	< 0.001	1.5	1	0.22	0.0076
FreqD X FreqA	12.4	1	< 0.001	12.9	1	< 0.001	0.0007
FreqA X FreqBC	15.7	1	< 0.001	16.5	1	< 0.001	0.0009
PC[FreqABC / LogitABC] X PhraseABCD	10.8	1	< 0.001	16.5	1	< 0.001	0.0001
LogitBCD X PC[FreqABC / LogitABC]	4.4	1	0.04	5.1	1	0.02	0.0002
FreqABCD X Length	7.3	1	0.01	8.8	1	0.003	0.0004
FreqABCD X FreqAB	10.2	1	< 0.001	4.5	1	0.03	0.0006
LogitABCD X PC[FreqCD / LogitCD]	4.2	1	0.04	4.6	1	0.03	0.0002

Notes. Denominator *df* = 4297. LLRT stands for log-likelihood ratio test.

Table 6: *Results of the Production Duration Analysis – Part 2*

Variable	MCMC β	95% CI	*t*	MCMC *p*
Number Syllables	0.0929	0.0818 to 0.1035	13.3	< 0.001
Manner of articulation	--	--	--	--
FreqB X PhraseABCD	0.0109	0.0026 to 0.0186	2.13	0.01
FreqC	-0.0046	-0.0108 to 0.0017	-1.2	0.15
FreqD X FreqA	-0.007	-0.0102 to -0.0039	-3.5	< 0.001
FreqA X FreqBC	0.0119	0.0073 to 0.0167	3.9	< 0.001
PC[FreqABC / LogitABC] X PhraseABCD	0.029	0.0182 to 0.0413	3.9	< 0.001
LogitBCD X PC[FreqABC / LogitABC]	-0.0063	-0.011 to -0.0021	-2.2	< 0.001
FreqABCD X Length	0.0052	0.0023 to 0.008	2.9	< 0.001
FreqABCD X FreqAB	0.0141	0.0086 to 0.0196	*4*	< 0.001
LogitABCD X PC[FreqCD / LogitCD]	-0.0072	-0.0131 to -0.0021	-2.1	0.01

Notes. MCMC stands for Markov Chain Monte Carlo. *CI* stands for confidence intervals.

Table 7: *Percentage of Deviance Explained by Family of Probabilistic Measures*

	Onset Latency	Production Duration
Frequency of Occurrence	0.21%	1.11%
Log Probability of Occurrence	0.93%	0.005%
Mutual Information	0.85%	0.00%

Table 8: *Percentage of Deviance Explained by N-gram*

N-Gram	Onset Latency	Production Duration
1	0.25%	0.96%
2	0.09%	0.17%
3	0.29%	0.03%
4	0.06%	0.12%

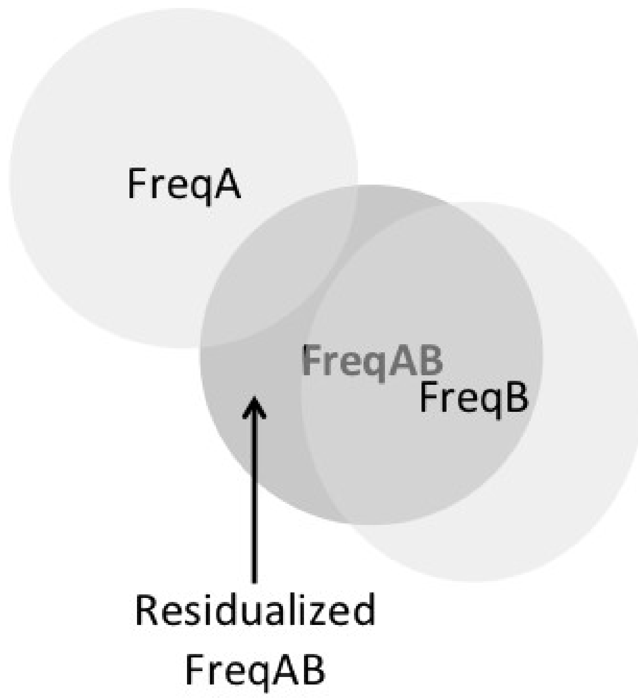
Figure Captions

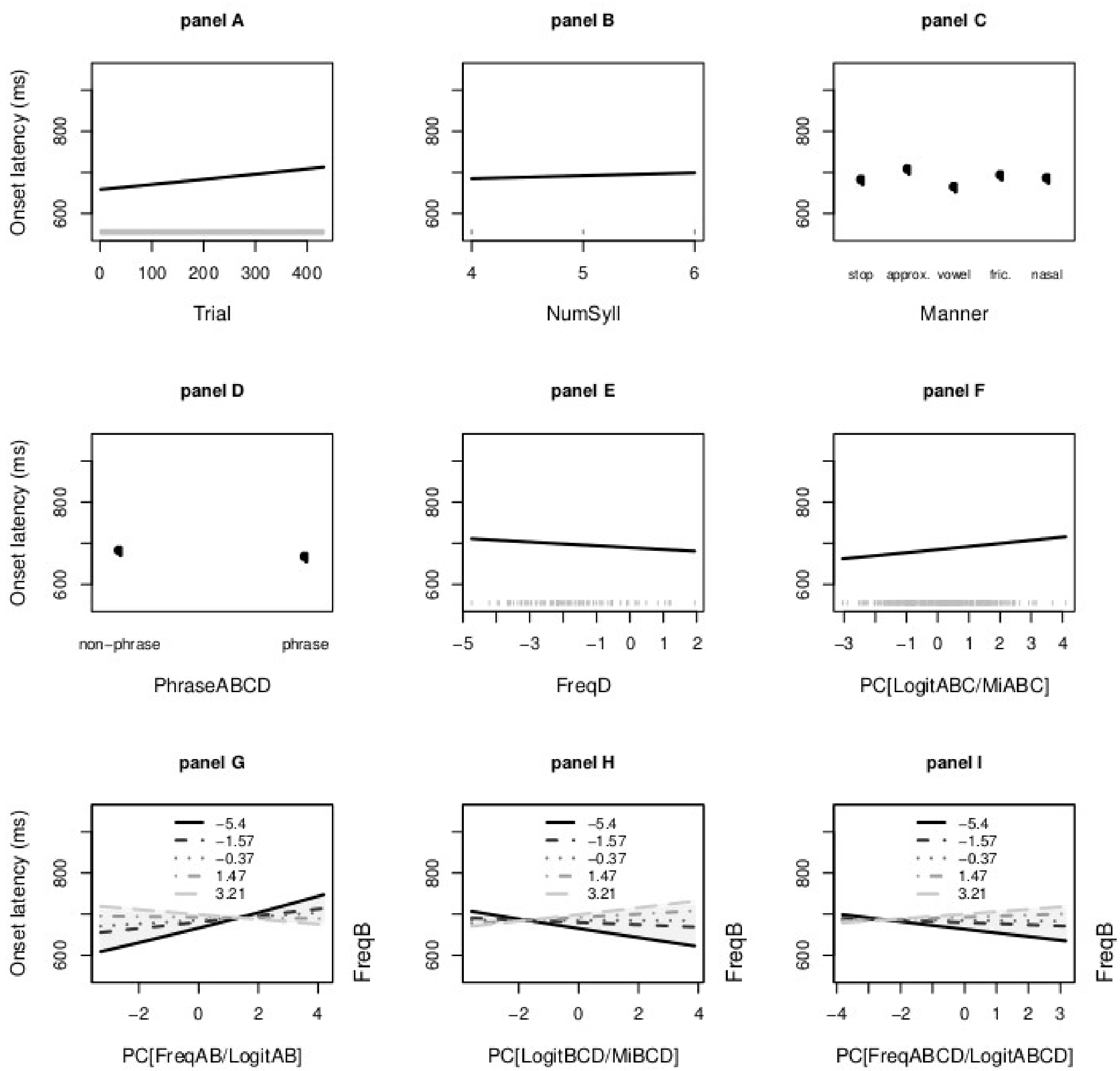
Figure 1. Residualization of FreqAB.

Figure 2. Onset latency analysis results.

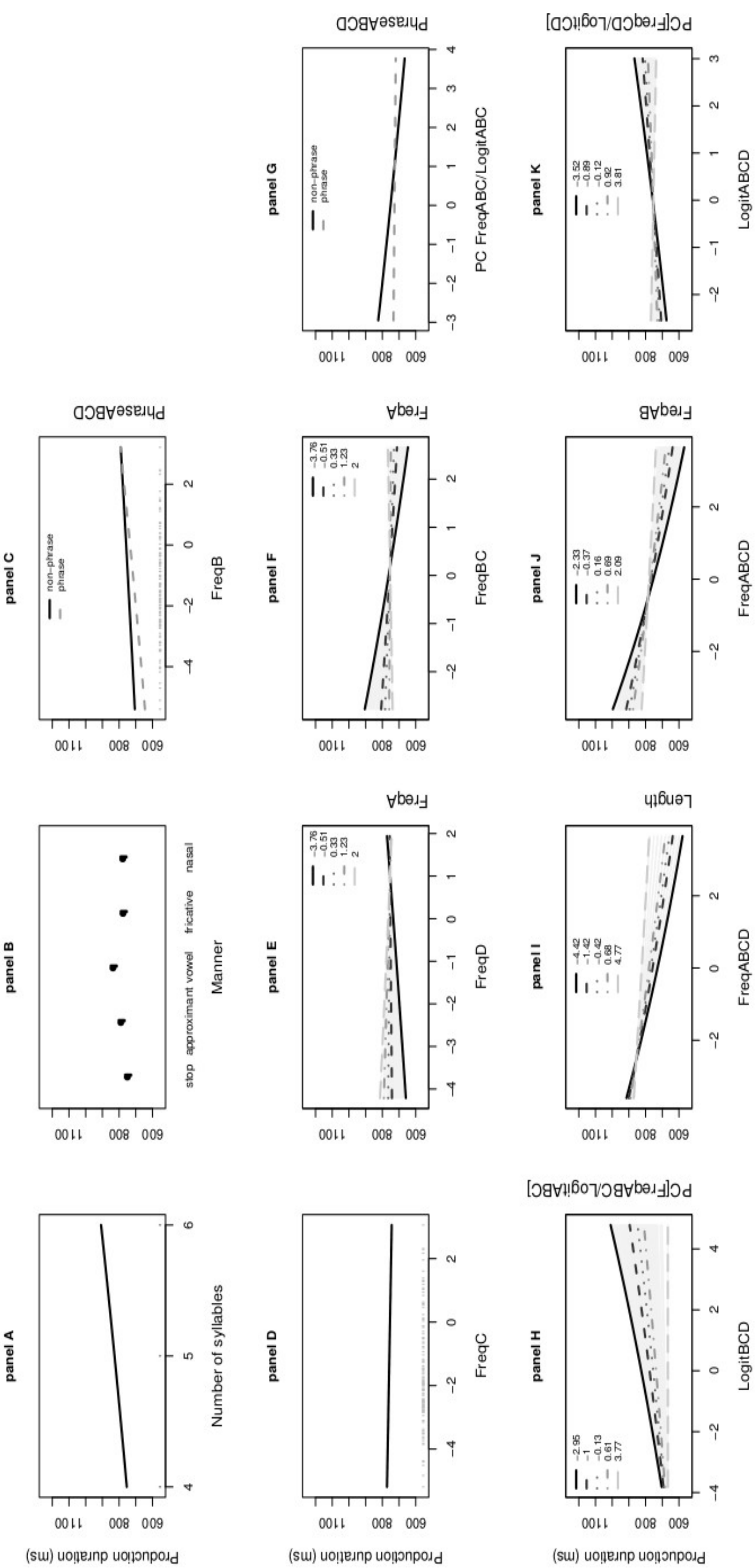
Figure 3. Production duration analysis results.

Figures





panel A
Onset latency (ms)
Trial
panel B
NumSyll
panel C
stop approx. vowel fric. nasal
Manner
panel D
non−phrase phrase
PhraseABCD
panel E
FreqD
panel F
PC[LogitABC/MiABC]
panel G
−5.4 −1.57 −0.37 1.47 3.21
FreqB
PC[FreqAB/LogitAB]
panel H
PC[LogitBCD/MiBCD]
panel I
PC[FreqABCD/LogitABCD]


panel A
panel B
panel C
panel D
panel E
panel F
panel G
panel H
panel I
panel J
panel K
Production duration (ms)
Number of syllables
Manner
stop approximant vowel fricative nasal
FreqB
PhraseABCD
non-phrase
phrase
FreqC
FreqD
FreqA
FreqBC
PC FreqABC/LogitABC
LogitBCD
PC[FreqABC/LogitABC]
FreqABCD
Length
FreqAB
LogitABCD
PC[FreqCD/LogitCD]